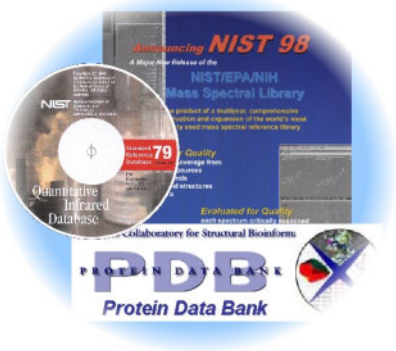
9. Chemical and Biochemical Data



Data acquisition, manipulation and handling are mandates for NIST activities. Thus, CSTL activities in the area of chemical and biochemical data are numerous and broad. These activities can be subdivided beyond chemical or biochemical activities, into areas of experimental data and derived data, which are obtained from modeling processes or statistical analysis of other data. Some of the data included in the databases are rigorously measured and evaluated, and thus can be called NIST Standard Reference Databases (SRD). Other databases, due to the nature and source of their data, are less rigorously evaluated, and therefore not termed a reference database. Both classes of NIST databases have comparable utility and significance, because they contain the most accurate data in the country, if not the world.

CSTL focuses on the areas of highest impact across industries, and draws on its expertise in data handling, and its portfolio of chemical and biochemical activities.

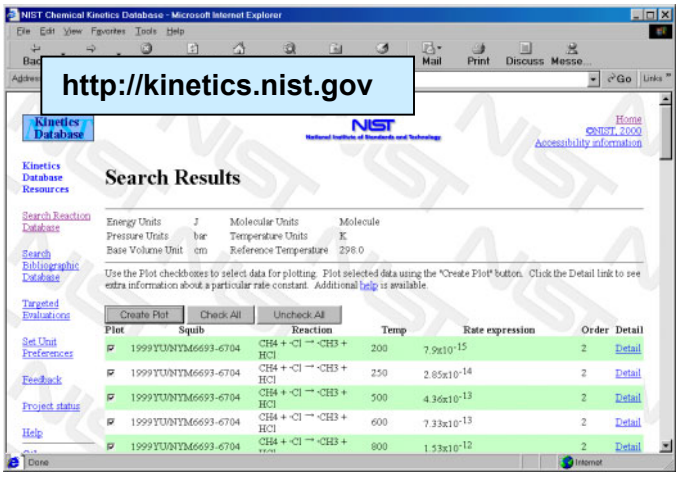
Kinetics Database on the Web

T.C. Allison, R.D. Levin, R.E. Huie (838), and C.-Y. Lin (Contractor)

The NIST Chemical Kinetics Database is the largest collection of chemical kinetics data available and is an extremely valuable resource for researchers in a number of scientific fields, including combustion modeling, atmospheric chemistry, and chemical vapor deposition. In fact, the reaction rates of chemical reactions are of fundamental importance in all chemical systems.

Kinetic data are needed by researchers to understand each phase of a substance's lifetime, from the design stage of industrial production processes, to determining effects on the global atmosphere, to devising appropriate disposal methodologies.

The NIST Chemical Kinetics Database has now been updated and made available via the World Wide Web and gives users powerful searching capabilities, more options for unit conversions, a tabular display of rate constant data, and a plotting interface. The database is linked to the NIST Chemistry WebBook wherever possible. Users now have the ability to access Targeted Evaluations that contain recommended values for kinetic and thermodynamic quantities obtained via extensive literature review.



New enhancements to the database will concentrate on the delivery of data in useful formats and on interfacing with other databases and projects. In particular, we will add the ability to import and export data in an XML-based format.

The Targeted Evaluations area also contains a JANAF-type display of thermodynamic data derived from *ab initio* calculations. These data are also available in the NASA polynomial format. All of these features are available through a simple and intuitive web interface.

Computational Chemistry Comparison and Benchmark Database

R. D. Johnson III (838)

The CCCBDB provides a benchmark set of molecules for the development and evaluation of new theoretical methods, providing users with a readily accessible means of comparing existing computational chemistry methods with experiment and analyzing the magnitude of error in computed thermochemical properties associated with time-saving approximations. The initial focus of the database was exclusively thermochemistry and structure, but has recently been expanded to include other properties of interest to the chemical industry, such as transition state structures and energies for kinetics, dipole moments, and ionization energies. The CCCBDB contains a set of 615 benchmark molecules with reliable thermochemical and spectral data for which both the experimental values and uncertainties have been evaluated. We are generating data from *ab initio* calculations for comparison with each other and with experiment. New data in the latest release include heat capacities, bond descriptions for isodesmic reaction predictions, electric dipole moments, ionization energies, and barriers to internal rotation (which are needed for the accurate prediction of entropies). Computed data include all of the above as well as atomic charges, HOMO-LUMO gaps, and diagnostics for the calculations. The calculations cover 25 different theoretical methods and seventeen basis sets. To make the database more useful and educational to non-expert users of computational chemistry, a series of tutorials and easy-to-understand graphic visualizations of the comparative results have been implemented. These help the user understand the inherent magnitudes and sources of error in different calculations. About 4,000 web pages are accessed every month.

The Computational Chemistry Comparison and Benchmark Database (CCCBDB) is designed to answer the question "How good is that quantum chemistry calculation for a given property?"

The CCCBDB website is designed to enable the non-expert academic or industrial user to choose the optimum method available to calculate properties of interest, and to understand the inherent magnitude and sources of error in the calculations.

http://srdata.nist.gov/cccbdb

The database contains the results from over 50,000 calculations (up from 30,000 last year) on 615 molecules of industrial interest for which good thermal and spectral experimental data are available.

Software for Predicting Chemical Reactivity

K.K. Irikura (838)

Predicting how molecules will react is one of the most important tasks required of chemists. It helps avoid unwanted side-reactions, design molecules with desired properties, and control thermal or oxidative decomposition.

Traditionally, predictions are obtained entirely through a chemist's intuition. Although this procedure is nearly instantaneous, obviously, it is error-prone and limited by the chemist's knowledge and skill.

Subsequent chemical R&D, which is typically expensive, depends upon the accuracy of the intuitive predictions. When intuition falls short, the R&D fails and the entire process must be repeated, resulting in costly losses in productivity. "Isopotential searching" (IPS) is a computational procedure in which the atoms of a molecule are moved semi-randomly.

To supplement human chemical intuition and thus improve the efficiency of R&D efforts, new software has been written for predicting and discovering chemical reactions. It employs the "isopotential searching" (IPS) method developed recently at NIST.

Researchers in CSTL have used this software to identify some new decomposition mechanisms for the military explosive known as RDX, predicting that NO_3 and/or N_2O_5 are reaction products. This suggests new strategies for detecting concealed quantities of the compound.

Beta and general releases of the software are planned for FY2002.

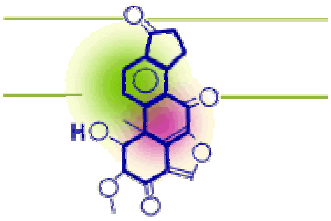
The feasibility of the atomic motion is computed quantum mechanically, and the atoms are moved again. The process is repeated until a reaction is discovered. Only those motions are tested that are of approximately equal energy, that is, all tested atomic configurations lie on an isopotential contour. This restriction leads to better efficiency than in conventional simulations. The purpose of the new software is to transfer the IPS technology to researchers outside NIST for use in diverse applications such as reaction design, catalyst screening, mass spectrometry, and reactive hazards evaluation.

The NIST Mass Spectral Database: Extending the Evaluation

S.E. Stein, A.I. Mikaia, J.K. Klassen, D.V. Tchekhovskoi, C.L. Clifton, Q.-L. Pu, W.G. Mallard (838), D. Zhu, and Y. Mirokhin (Contractors)

The NIST MS database is the most widely distributed mass spectra database in the world. Annual sales are over 2500 units, which represents about one-half of the GC/MS instruments sold worldwide. The database and its associated fully tested software are increasingly becoming the standard for industrial use.

The determination of the identity of a compound is a central problem in chemistry. For volatile substances, the most widely used, sensitive, and definitive "fingerprint" for making such identifications is the electron-ionization mass spectrum. In practice, identifications begin with automated mass spectral "library searching" against a comprehensive library of reference spectra. The reliability of such identifications depends directly on both the quality of the reference library and the algorithms for matching mass spectra. The goal of this program is to provide a fully evaluated mass spectral database with tested and documented search algorithms to enable the reliable identification of unknown organic compounds by gas chromatography/mass spectrometry (GC/MS). The existing NIST Mass Spectral Database is the most widely used database for GC/MS analysis. Over the past year, critical evaluation has been completed for all spectra received since the previous release. New data for retention indices of a large number of chemicals are being added to ensure that the NIST MS Database continues to fulfill the needs of the organic analytical community. A significant effort has also been devoted to improving the consistency of chemical names and chemical structures. The accuracy of Chemical Abstracts registry numbers has been thoroughly established with the assistance of the Chemical Abstracts Service. At the same time, work on new algorithms for searching of MS/MS spectra has been coupled with new efforts at determining MS/MS spectra for a variety of well-known conditions in order to test the algorithms. The extension of the algorithms to prediction of MS-MS results will allow the electron impact database, by far the largest set of data in mass spectrometry, to be used by new, alternative ionization techniques. This will assist data analysis in the rapidly advancing field of MS-MS.



The NIST Mass Spectral library is used in drug discovery screening, drug testing, forensic analysis, pesticide and pesticide residue testing, as well as in numerous industrial quality control and product development labs.

Automated Gas Chromatography/Mass Spectral Deconvolution and Analysis: Tools for Automating and Improving the Use of GC/MS Instruments (AMDIS)

S.E. Stein, J.J. Reed, W.G. Mallard (838), and O. Toropov (Contractor)

NIST has developed a tool for automatically analyzing gas chromatographic - mass spectrometric (GC/MS) data files: Automatic Mass Spectral Deconvolution and Identification Software (AMDIS). The AMDIS package has been widely adopted by GC/MS instrument companies.

This program was originally supported by the Defense Threat Reduction Agency (DTRA) to provide a method for determining whether chemical weapons banned under the CWC are present in samples analyzed by GC/MS. In order to ensure that any proprietary or national security information not be revealed during the analysis, this had to be implemented without the operator examining the data.

AMDIS provides data deconvolution combined with noise analysis resulting in a tool that is free of operator bias and is capable of providing more sensitive detection, and higher resolution than previously possible. Although initially designed for the DTRA, AMDIS has subsequently been widely distributed by instrument manufacturers to aid their users in analyzing GC/MS data from complex mixtures

Compared with manual analysis, the AMDIS software allows the data to be analyzed more quickly and without operator bias. In addition, for a complex chromatographic region the overlapping components cannot be properly treated using manual "background subtraction."

Direct support to U.S. treaty efforts are supplied by NIST through this work. These include evaluation of data for use in inspections, technical assistance to the U.S. Delegation to the Organization for the Prohibition of Chemical Weapons (OPCW), and direct assistance to U.S. inspection personnel. AMDIS is currently in use for on-site inspections by the OPCW, and it is increasingly being accepted by the member states of the OPCW as the reference standard for their own work.

The NIST WebBook: - Chemical Reference Data for Industry

W.G. Mallard, P.J. Linstrom, D.H. Frizzell (838), and J.F. Liebman (UMBC)

One goal of this project is to locate, transcribe, and evaluate the enormous amount of published thermochemical data, such as heats of formation, entropies, heat capacities, and heats of reaction; as well as thermophysical property data, such as vapor pressure, viscosity, boiling points, melting points, etc.

The NIST Chemistry WebBook continues to be one of the most widely used resources for chemical and physical property data in the world. Usage has continued to grow at about 20% per year with over 350,000 distinct IP addresses accessing the WebBook in the last year.

During FY01 the seventh edition of the NIST Chemistry WebBook was released and properties for more than 40,000 compounds are available. Extensive additional vapor pressure data originating from the Thermodynamics Research Center have been added, in addition to new high-precision data for industrially important fluids, and new calculational modules to allow transport properties from the equations of state for the fluids.

Often these data are effectively unavailable to the technical community because of the inherent slowness of locating and assessing the data. A second aspect is to make this and other data such as infrared (IR), ultraviolet (UV), and mass spectra available and easy to access.

http://WebBook.nist.gov

Noteworthy Facts about the NIST WebBook:
During FY01, the total number of distinct internet addresses that accessed the WebBook was in excess of 350,000. Growth in usage has been fairly steady from year to year – with same-period usage increasing by about 20% in each of the last four years. There are between 10,000 and 20,000 users per week, from industry, government, and academia.

Resources of the WebBook are being directly accessed by other computer programs used in problem-solving desktop environments by chemical engineers from the Colorado School of Mines, Gaussian Inc., Exxon-Mobil, DuPont Chemical, and Reaction Design, Inc., and other academic partners.

Thermodynamics Research Center (TRC) Comprehensive Program on Critical Data Evaluation

M.L. Frenkel, R D. Chirico, G.R. Hardin, R.A. Stevenson (838), Q. Dong, X. Yan, X. Hong (Contractors), and V. V. Diky (Belarussian State U., Belarus)

Details provided in the **Physical Property Data** section.

Development of Efficient Tools for Computational Kinetics

C.A. Gonzalez (838)

The main goal of this project is the generation of computational tools that will allow scientists and engineers to perform reliable studies of the kinetics governing gas-phase chemical reactions relevant to industrial and atmospheric processes.

The use of modern computational chemistry methodologies in the prediction of molecular properties has become increasingly popular mainly due to significant improvements in the algorithms and the advent of powerful computer resources. This is particularly true in the area of thermochemistry, where researchers in industry and academia perform quantum chemistry calculations on a routine basis. Despite this progress, the use of similar methodologies in computational kinetics has been slow. In fact, computational kinetics remains an obscure area being used most of the time by experts in the field. There is no doubt that the application of quantum chemistry calculations in the area of computational kinetics will have significant impact. In order for this to happen, however, it is critical that state-of-the-art methodologies be widely available to the scientific community. In this project, a series of modules that compute rate constants as a function of the temperature have been implemented into a generic piece of software. Given its wide applicability, Canonical Transition State Theory (CTST) has been chosen as the standard method to compute rate constants. Tunneling corrections are calculated using the one-dimensional Wigner, Symmetrical Eckart, or Unsymmetrical Eckart approaches. A module that treats normal modes as hindered rotors has also been implemented. In addition, properties such as vibrational frequencies, heats of reaction, entropies, and partition functions are also computed. The program has been implemented in standard FORTRAN 77, and it has the capability of reading output files from one of the most popular quantum chemistry programs.

A series of modules that compute rate constants as a function of the temperature have been implemented into a generic piece of software. This will allow quantum chemistry calculations to be more easily applied in the in the area of computational kinetics. Beta-test versions of the program running on PCs and UNIX workstations have been distributed among scientists inside and outside NIST.

Fifth International Conference on Chemical Kinetics

J.W. Hudgens, and R.E. Huie (838)

The Fifth International Conference on Chemical Kinetics was held in early summer 2001 for scientists that have a common interest in chemical kinetics, both in the gas and liquid

The utilization of chemical kinetics is predominantly in the modeling of complex chemical processes. Important examples include the models of the stratosphere that underlie the phase-out of halocarbons, models of urban air pollution, and models of chemical vapor-deposition processes. Kinetic models are now being developed to allow the modeling of soot formation and, ultimately, its control. The conferences in this series are designed not to focus specifically on a single problem to which chemical kinetics is applied. Rather, the goal was to have a much wider ranging conference where practitioners and users of chemical kinetics would be exposed to more diverse information, so that they could draw upon much more of the wealth of the field.

This broadly based conference has filled a major need in the kinetics community, of both those who use and those who generate chemical kinetics data. It has allowed cross-fertilization across the field and the integration of results from various sub-disciplines.

http://www.cstl.nist.gov/div838/kinetics2001/

Exploring Reaction Kinetics and Physical Properties of Ionic Liquid Systems

D. Behar, P. Neta, J. Magee, and M. Frenkel (838),and G. Kabo (Belarusian State University, Belarus)

Details provided in the **Environmental Measurements** section.

Kinetics of Reactions of the Hydroxyl Radical

V.L. Orkin, M.J. Kurylo, and R.E. Huie (838)

Details provided in the **Environmental Measurements** section.

Thermochemical and Chemical Kinetics Data for Organometallic Compounds

D.R. Burgess, Jr., J.W. Hudgens, and E.P. Hunter (838)

Organometallic compounds are used during manufacturing processes to deposit metals in semiconductor, optical, fuel cell, MEMS, and NEMS devices.

The thermochemical properties and reaction kinetics of most useful organometallic compounds and related molecular precursors are poorly characterized. CSTL scientists are working to obtain these properties through specialized activities that involve both theoretical estimations and laboratory measurements. CSTL, in cooperation with the Standard Reference Data Program, compiles and evaluates currently available thermochemical data of organometallic compounds and related precursors. This data will become available through an external NIST website. The second activity supplements available data by using *ab initio* and semi-empirical calculations to develop reaction mechanisms from computed molecular structures, thermodynamic properties, and spectroscopic properties of Group III and Group V compounds. The third activity is a laboratory measurement program that measures the thermal decomposition kinetics of organometallic compounds. A turbulent/laminar flow apparatus for thermodecomposition kinetic rate measurements of organometallic compounds was designed and constructed. The kinetic apparatus resides in a

By using advanced measurement and computational facilities, CSTL researchers will reliably measure, estimate, and disseminate fundamental thermochemical and chemical kinetic properties of organometallic compounds.

standing hood facility and consists of a heated flow reactor, flow controllers, and a Fourier-transform infrared spectrometer. Each measurement series produces Arrhenius-like equations describing homogeneous and heterogeneous decomposition rates as a function of temperature.

Thermal Decomposition of Chlorinated Hydrocarbons

J.A. Manion, D.R. Burgess, Jr., W. Tsang (838), and I.A. Awan (U. of Peshawar, Pakistan)

Chlorinated hydrocarbons are widely utilized throughout the chemical industry, both as end products and as precursors for a wide variety of useful products, including plastics, solvents, pesticides, refrigerants, and other products.

Attempts to understand and model the chemistry associated with the production, disposal, and environmental fates of chlorinated materials require that reliable kinetic and thermodynamic values are available for these compounds. Recent evaluations have brought to light some existing problems with both kinetic and thermodynamic data, and we have suggested some significant changes to currently available values. These changes have important implications regarding the chemistry of these compounds. Additionally, a major current thrust is to develop computational methodologies enabling the rapid prediction of key physical properties. Species of interest frequently have highly polar substituents, including chlorine, and thermal decomposition of these compounds have proved to be among the most difficult to calculate accurately. It is therefore extremely important to have a reliable database with which to compare alternate predictive strategies. We have measured the rate constants for the decomposition reactions of a number of chlorinated C2 and C3 hydrocarbons at temperatures near 1000 K using single-pulse shock tube studies. These experiments have tested and confirmed our recently proposed changes to heats of formation of highly chlorinated compounds, as well as resolved contradictory information on the kinetics of these compounds. *Ab initio* computational studies have also been carried out to extend the range of compounds for which reliable information is available.

The information provides computational chemists with crucial benchmark data on highly polar compounds, which are key species in a much wider variety of industrial chemistries.

SpectroML- An Extensible Markup Language for Molecular Spectroscopy Data

M. A. Rühl, M.G. Peschke, (U. of Applied Sciences, Germany)
J.C. Travis, G.W. Kramer (839), and R. Schäfer (U. of Applied Sciences, Germany)

Today's analytical chemists need not only to interchange data generated within a specific technique, they need to interchange, import, export, store, and combine all their data from multiple sources at multiple sites. The interchange and storage of analytical chemistry data has long been hampered by multiple, incompatible data formats.

In the last few years, the emergence of platform and application independence through the Internet has advanced the possibilities for data interchange. SpectroML was created initially to serve as a model implementation of a markup language for molecular spectrometry and ultimately to provide a web-based mechanism for interchanging UV/visible spectral data generated on different spectrophotometers among ourselves, our colleagues at other National Metrology Institutes, and our NIST-Traceable Reference Material (NTRM) vendors.

Over the past year, CSTL researchers have developed a "web-aware" mechanism for instrument-to-instrument, instrument-to-application, and application-to-application data interchange called SpectroML: an extensible markup language for molecular spectrometry data.

We created SpectroML in the Extensible Markup Language (XML) and defined and documented it with both a DTD (Document Type Definition) and a schema. We have developed a few demonstration applications, applets, and stylesheets (used as templates to map document type conversions). SpectroML will evolve as the ASTM task group refines and expands it to work across the breadth of molecular spectrometry and perhaps even beyond to atomic spectrometry and chromatography.

Representatives from several major instrument companies and the *ASTM E13 Molecular Spectrometry Committee* support a markup language for molecular spectrometry based on SpectroML. NIST has developed demonstration applications but the impact of SpectroML will expand and evolve as an ASTM E13 task group and others refine and implement its use to applications across the breadth of molecular spectrometry and perhaps to other techniques.

An Amino Acid Exchangeability Measure Derived from Experimental Data

A.B. Stoltzfus (831), and L.Y. Yampolsky (CARB/UMBI)

A new statistical model of the mean effect EX_{ij} of changing amino acid i to amino acid j on protein activity has been developed. The new model is based on data from 12 experimental laboratory studies in which amino acid exchanges were engineered systematically.

The aim of this work is to derive useful information on the mean effects of amino acid exchanges from the results of experimental exchanges of amino acids under laboratory conditions. Measures of pairwise similarity between amino acids have many different applications in bioinformatics. The most commonly used measures, PAM and BLOSUM, summarize patterns of evolutionary divergence. When used to gauge the effects of amino acid change on proteins, however, evolution-based measures are problematic, because they reflect not only effects of amino acid changes on proteins but also the rates of underlying mutations necessary for evolutionary change to occur. Experimental data could be used to avoid mutational biases and thus clarify the mean effects of amino acid changes in proteins. The challenges are to find a sufficiently large set of data from experiments whose design does not impose strong biases, and to combine results from diverse studies. To this end, the results of 9334 single amino acid exchanges were tabulated from 12 studies in which amino acid exchanges were engineered systematically, without imposing substantial biases. Results of diverse studies were transformed to a common scale using a statistical model relating frequency to severity of effect on activity, based on 7 studies that provided information on the activity of variant proteins relative to the wild-type. The resulting estimation of the mean effect EX_{ij} of changing amino acid i to amino acid j was evaluated by a jack-knife procedure in which a subset of the studies is used to predict the results of the remaining studies, with the predictive power being analyzed by logistic regression and compared to the power of other predictors.

In most cases the EX model is a better predictor than existing evolutionary divergence methodologies and avoids problems due to mutational biases. The results have been presented in a seminar and have stimulated interest from a pharmaceutical company in the possibility of a CRADA to support further work.

As shown in the table, EX is a better predictor than all measures tested except for BLOSUM, even though most of these measures are based on far larger amounts of data from comparative sequence analysis (PAM, BLOSUM, Grantham) or from structural modeling (MJ).

	Combined surface & buried sites					*Separate sites*	
Predictor:	BLO	EX	PAM	MJ	1/Gra	EX	KG
Log Likelihood	-169.0	-147.7	-138.5	-91.4	-36.0	-395.9	-190.3

Separate EX matrices for surface and buried sites have been computed, and this pair of matrices outperforms a comparable pair of matrices based on evolutionary divergence (KG, Koshi-Goldstein). An asymmetric EX matrix, in which EX_{ij} and EX_{ji} are different, can also be computed: this matrix outperforms all other predictors tested, including BLOSUM.

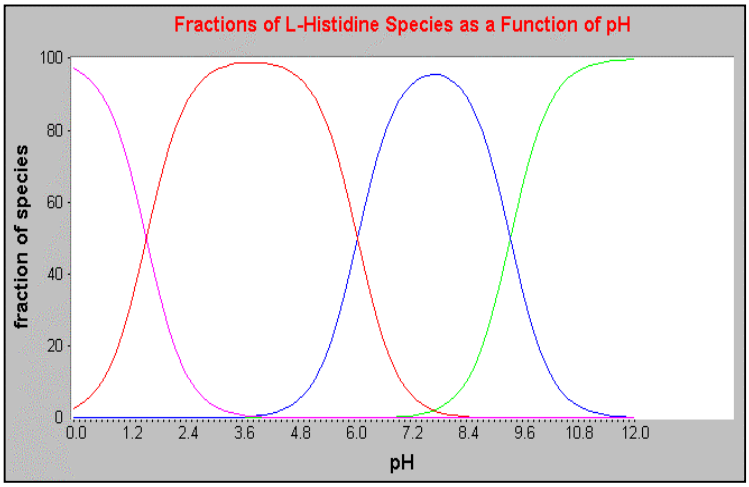
Biothermodynamics

R.N. Goldberg and Y. Tewari (831)

The International Union of Pure and Applied Chemistry (IUPAC) Commissions on Biophysical Chemistry and Thermodynamics identified the need for a comprehensive, critically evaluated tabulation of thermodynamic quantities for the ionization reactions of biological buffers as a critical need in biochemistry.

A comprehensive tabulation of thermodynamic quantities for the ionization reactions of biological buffers does not yet exist. However, these buffers are essential components of every biological preparation ranging from nucleic acids through to stem cells and whole tissues.

CSTL researchers completed a review article that contains selected values of thermodynamic quantities for the ionization reactions of 64 buffers, which are used in biological research. Since the aim is to be able to predict values of the ionization constant at temperatures not too far from ambient, the thermodynamic quantities which are tabulated are the pK, standard molar Gibbs energy $\Delta_r G°$, standard molar enthalpy $\Delta_r H°$, and standard molar heat-capacity change $\Delta_r C_p°$ for each of the ionization reactions at 298.15 K. The selection of thermodynamic data for each buffer is discussed. A summary data table is available for calculating fractions of the various buffer species present in solution as a function of pH. An example result for L-histidine is shown.



The amino acid L-histidine can form four distinct ionic species as a result of changing pH. Using values of thermodynamic properties assembled in the new Biological Buffer Database the figure shows the relative fraction of each species as a function of pH:

Purple – L-histidine^{2+},
Brown – L-histidine^{+},
Blue – L-histidine^{0},
Green – L-histidine^{-}

Accurate Measurement of the Half-life of ^{76}As

R.M. Lindstrom (839), M. Blaauw (Delft U, The Netherlands), and R. F. Fleming (U of Michigan)

Knowledge of half-lives for radioactive decay is necessary for all applications of radioactivity. In the course of high-accuracy measurements of arsenic, we found that the most recently published and compiled half-life of ^{76}As did not fit our data as well as the earlier accepted value.

To redetermine the half-life of ^{76}As, five ^{76}As sources were measured on four Ge detector systems, and the decay data fit to an exponential function by two different nonlinear least-squares methods. We determined $t_{1/2}$ = 1.0938 d with an expanded uncertainty U = 0.0009 d. This result is 1.5% higher than the recent value, and in agreement with the older, less precise, consensus value. To verify our methods, we also measured the half-life of ^{198}Au, one of the best-studied nuclides because of its fundamental importance in the unification of the energy and wavelength in photon spectroscopy. In this case our number agrees with the most precise (and adopted) value to within 0.1%. In the course of our work we have identified a weakness in the older data fitting procedure and have developed a new algorithm for combining data that is superior to the one previously in use. A detailed publication on the NIST-IRI methodology is in preparation. Now that the power of the new NIST methodology has been established, a continuing low-level effort to redetermine a selected set of half-lives will be carried out, in collaboration with researchers in the Physics Laboratory.

Impact:
An erroneous published value for a physical parameter has been corrected. This has led to a revision of the internationally accepted value of the ^{76}As half-life, and has led to the establishment of a new algorithm for combining data that is superior to the one previously in use.

Thermodynamic Properties for Environmental Fate and Risk Evaluation: Arsenic and Chromium

D.R. Kirklin and D.G. Archer (838)

Details provided in the **Environmental Measurements** section.

Standards for Calorimetry and Thermodynamics

D.R. Kirklin and D. G. Archer (838)

Certification measurements for two new SRMs for calibration of DSCs that operate at subambient to ambient temperatures were made with newly designed and constructed apparatus. Certification of a renewal lot of SRM 217 by combustion calorimetry was also conducted this year.

Thermodynamic and calorimetric measurements have wide usage, ranging from product specification and quality control to research. Many calorimetric or thermodynamic instruments must be calibrated and/or validated for proper operation. Standard Reference Materials and other highly accurate property data for calibration and/or validation of calorimetry and other thermodynamic measurements have been developed. A description of our previous high-accuracy, enthalpy-of-solution measurements for sodium chloride has now been completed and accepted for publication. These measurements establish this material as a convenient reference material for solution calorimeters. At the request of parties from PTB, IUPAC, and GEFTA, a guide for the practice of conversion of thermodynamic properties from one international temperature scale basis to another was prepared. This guide provides analysis and revision of the thermodynamic properties for two of NIST's calorimetric SRMs: SRM 715, polystyrene, and SRM 781, molybdenum.

Physikalisch-Technische Bundesanstalt

New Databases for Surface Analysis by Auger-Electron Spectroscopy and X-Ray Photoelectron Spectroscopy

C.J. Powell (837), A. Y. Lee, D. M. Blakeslee, H. Thai, J. R. Rumble, Jr. (TS)
A. Naumkin, A. Kraut-Vass, A. Jablonski, and W. S. M. Werner (Contractors)

NIST DATA

During the past year, a new version of the popular NIST X-Ray Photoelectron Spectroscopy Database (SRD 20) was released, the new NIST Electron Effective-Attenuation-Length Database (SRD 82) was released, and work advanced on a new NIST database for Quantification of Electron Spectroscopic Techniques (QUEST). Details are provided below.

- ***SRD 20*** provides identification of unknown spectral lines in user measurements, retrieval of data for selected elements, retrieval of data for selected compounds, and retrieval of data by scientific citation. From January to July 2001, there were approximately 30,000 requests per month for information from this database.
- ***SRD 82*** provides values of electron effective attenuation lengths and related data forAuger-electron spectroscopy (AES) and X-Ray photoelectron spectroscopy (XPS). Values of effective attenuation lengths (EALs) are needed mainly for measurements of thicknesses of overlayer films and to a much lesser extent for measurements of the depths of thin marker layers. SRD 82 was released in September 2001.
- In addition, work is well advanced on the new **QUEST** database to be used for AES and XPS analyses of materials with complex morphologies. This database will access data from SRD 64 and SRD 71 and, with additional data, will enable comparisons to be made of measured and simulated spectra for particular specimen morphologies and specified analytical conditions.

Protein Data Bank (PDB)

T.N. Bhat and G.L. Gilliland (831)

RCSB

The Research Collaboratory for Structural Bioinformatics (RCSB - members are Rutgers, San Diego Computing Center, and NIST) has been fully responsible for the management of the Protein Data Bank since July 1, 1999. The RCSB has provided new integrated systems for collecting, distributing and querying the contents of the PDB. These systems were designed with the expectation that there would be a change in the quantity, quality, and content of new structures. It is the goal of the PDB to enable new science through effective use of structure data. The biotechnology sector has generated vast amount of data and will continue to do so in the future. Consistent schema, uniform validation tools, and standard database interfaces are needed to allow efficient queries and distribution of these data. Quality and uniformity are two major issues for dependable and reliable results. The PDB Web sites provide users with direct query and reporting capabilities using the underlying databases. The query capabilities are quite extensive, and have been substantially improved with the introduction of the Molecular Information Agent (MIA), which provides frequently updated links to a growing number of databases. Querying across the complete PDB has nevertheless been limited by missing, erroneous, and inconsistently reported experimental data, nomenclature, and functional annotation. Inconsistency, in particular, reflects the evolution of experimental methods, functional knowledge of proteins, and methods used to

The PDB is an international repository for the processing and distribution of 3-D structure data of biological macromolecules determined experimentally by NMR and X-ray crystallography

PDB
PROTEIN DATA BANK

http://pdb.nist.gov

M.I.A.
Molecular Information Agent

http://mia.sdsc.edu/

process these data over the years. The result is that only PDB ID searches provide completely reliable results. The purpose of the data uniformity project that is underway is to address the non-uniformity in PDB data and thus enabling better query results for the customers. The first stage of data uniformity work on approximately 15000 legacy PDB entries is completed and the results are now distributed to users. These data uniformity efforts focused on a broad range of issues like data quality, data standards, and data exchange. To further enhance the interoperability of the structural data, CORBA standard is proposed and established (OMG Document Number: lifesci/2000-11-01) by the PDB. Our efforts on data uniformity and data standards are an on-going, and long-range program.

The total number of entries held by PDB is expected to double in the next three years and this increase is expected to further underscore the importance of data uniformity.

Bioinformatics Software Resource

T.N. Bhat and G.L. Gilliland (831)

Data standards and software tools are the two key elements of information technology. The proposed software resource is expected to support and catalyze such activities. In the area of bioinformatics and macromolecular structures extensive data uniformity efforts are underway. (Reported by Bhat et. al., in Nucleic Acid Research, 2001, Vol. 29, No. 1).

Up until now, most of the efforts in uniformity and standards in structural biology have focused on data. Recent developments in technology have created an urgent need for similar efforts in software standards as well.

The approach taken to provide this software resource involves:

a) development of a website that utilizes a database of web links and library of software and benchmark data for bioinformatics/computational Biology with particular emphasis on database development and data mining, and

b) development of standard application program interface (API) for many of the commonly used tasks in bioinformatics.

Further, a modern resource for data archival and distribution has also been established for the PBD described above. However, there is an urgent need for similar collective effort from the point of view of software and tools for bioinformatics and for macromolecular structural research to facilitate interoperability. Format conversions, standard object definitions and CORBA standards are some of the tasks of interest for API development. For this purpose a Bioinformatics Software Resource (BISR) has been developed. At present BISR provides key-word searchable web links for sites that distribute bioinformatics software. In the future BISR also plans to provide a stable archival and distribution facility for publicly available software in bioinformatics. Through this effort, our goal is to catalyze and contribute to the development of standard application program interface (API) for format conversions and data exchange.